DOCUMENT RESUME

ED 471 138 IR 021 861

AUTHOR	Kingsbury, Nancy
TITLE	Program Evaluation: Strategies for Assessing How Information Dissemination Contributes to Agency Goals. Report to Congressional Committees.
INSTITUTION	General Accounting Office, Washington, DC.
REPORT NO	GAO-02-923
PUB DATE	2002-09-00
NOTE	43p.
AVAILABLE FROM	U.S. General Accounting Office, P.O. Box 37050, Washington, DC 20013. Tel: 202-512-6000; Fax: 202-512-6061; Web site: http://www.gao.gov. For full text: http://www.gao.gov/new.items/ d02923.pdf/.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	Agencies; Evaluation Criteria; *Evaluation Methods; Federal Government; *Federal Programs; *Information Dissemination; Organizational Objectives; Program Effectiveness; *Program Evaluation; Program Improvement; *Public Agencies
IDENTIFIERS	Dwight D Eisenhower Professional Development Prog; Environmental Protection Agency; Expanded Food and Nutrition Education Program; *General Accounting Office

ABSTRACT

Federal agencies use information dissemination programs as one of several tools to achieve various social or environmental goals. Evaluations of five federal information dissemination program cases were examined to assist agency efforts in evaluating and improving the effectiveness of such programs. These cases were: Environmental Protection Agency (EPA) Compliance Assistance, the Eisenhower Professional Development Program, the Expanded Food and Nutrition Education Program (EFNEP), the National Tobacco Control Program, and the National Youth Anti-Drug Media Campaign. Useful evaluation strategies were identified that other agencies might adopt. This report discusses the strategies by which these five cases addressed their evaluation challenges. Contents include: the results in brief; background; scope and methodology; case descriptions; observations; and Agency comments. The following sections are highlighted: "Program Flexibility, Delayed Effects, and External Influences Posed Major Evaluation Challenges"; "Surveys and Logic Models Helped Address Most Challenges, but External Factors Supported These Evaluations"; and "Congressional Interest, Collaboration, Available Information and Expertise Supported These Evaluations." Other evaluation and guidance tools and related General Accounting Office (GAO) products are listed at the end of the document. Includes one table and four figures. (Contains 21 references.) (AEF)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 471 138

United States General Accounting Office

GAO

Report to Congressional Committees

September 2002

PROGRAM EVALUATION

Strategies for Assessing How Information Dissemination Contributes to Agency Goals

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

IR021861

G A O
Accountability * Integrity * Reliability

GAO-02-923

BEST COPY AVAILABLE

Contents

Letter 1

Results in Brief 2
Background 3
Scope and Methodology 6
Case Descriptions 7
Program Flexibility, Delayed Effects, and External Influences Posed Major Evaluation Challenges 9
Surveys and Logic Models Helped Address Most Challenges, but External Factors Were Rarely Addressed 13
Congressional Interest, Collaboration, Available Information and Expertise Supported These Evaluations 28
Observations 32
Agency Comments 33

Bibliography 35

Case Evaluations and Guidance 35
Other Evaluation Guidance and Tools 37

Related GAO Products 38

Table

Table 1: The Programs' Challenges and Their Strategies 13

Figures

Figure 1: Information Dissemination Program Logic Model 5
Figure 2: University of Wisconsin Cooperative Extension Logic Model 22
Figure 3: Logic Model for the National Youth Anti-Drug Media Campaign Evaluation 23
Figure 4: CDC Tobacco Use Prevention and Control Logic Model 27

Abbreviations

CDC	Centers for Disease Control and Prevention
CSREES	Cooperative State Research, Education, and Extension Service
EFNEP	Expanded Food and Nutrition Education Program
EPA	Environmental Protection Agency
ETS	environmental tobacco smoke
HHS	Department of Health and Human Services
NIDA	National Institute on Drug Abuse
NSPY	National Survey of Parents and Youth
OECA	Office of Enforcement and Compliance Assurance
OMB	Office of Management and Budget
ONDCP	Office of National Drug Control Policy
USDA	U.S. Department of Agriculture

GAO
Accountability • Integrity • Reliability

United States General Accounting Office
Washington, DC 20548

September 30, 2002

The Honorable Fred Thompson
Ranking Minority Member
Committee on Governmental Affairs
United States Senate

The Honorable Stephen Horn
Chairman
The Honorable Janice D. Schakowsky
Ranking Minority Member
Subcommittee on Government Efficiency, Financial
Management, and Intergovernmental Relations
Committee on Government Reform
House of Representatives

Federal agencies are increasingly expected to focus on achieving results and to demonstrate, in annual performance reports and budget requests, how their activities will help achieve agency or governmentwide goals. We have noted that agencies have had difficulty explaining in their performance reports how their programs and activities represent strategies for achieving their annual performance goals. Agencies use information dissemination programs as one of several tools to achieve various social or environmental goals. In programs in which agencies do not act directly to achieve their goals, but inform and persuade others to act to achieve a desired outcome, it would seem all the more important to assure decision makers that this strategy is credible and likely to succeed. Various agencies, however, fail to show how disseminating information has contributed, or will contribute, to achieving their outcome-oriented goals.

To assist agency efforts to evaluate and improve the effectiveness of such programs, we examined evaluations of five federal information dissemination program cases: Environmental Protection Agency (EPA) Compliance Assistance, the Eisenhower Professional Development Program, the Expanded Food and Nutrition Education Program (EFNEP), the National Tobacco Control Program, and the National Youth Anti-Drug Media Campaign. We identified useful evaluation strategies that other agencies might adopt. In this report, prepared under our own initiative, we discuss the strategies by which these five cases addressed their evaluation

challenges. We are addressing this report to you because of your interest in encouraging results-based management.

To identify the five cases, we reviewed agency and program documents and evaluation studies. We selected these five cases because of their diverse methods: two media campaigns were aimed at health outcomes, and three programs provided assistance or instruction aimed at environmental, educational, and health outcomes. We reviewed agency evaluation studies and guidance and interviewed agency officials to identify (1) the evaluation challenges these programs faced, (2) their evaluation strategies to address those challenges, and (3) the resources or circumstances that were important in conducting these evaluations.

Results in Brief

Assessing a program's impact or benefit is often difficult, but the dissemination programs we reviewed faced a number of evaluation challenges—either individually or in common. The breadth and flexibility of some of the programs made it difficult to measure national progress toward common goals. The programs had limited opportunity to see whether desired behavior changes occurred because change was expected after people made contact with the program, when they returned home or to work. Asking participants to report on their own attitude or behavior changes can produce false or misleading information. Most importantly, long-term environmental, health, or other social outcomes take time to develop, and it is difficult to isolate a program's effect from other influences.

The five programs we reviewed addressed these challenges with a variety of strategies, assessing program effects primarily on short-term and intermediate outcomes. Two flexible programs developed common measures to conduct nationwide evaluations; two others encouraged communities to tailor local evaluations to their own goals. Agencies conducted special surveys to identify audience reaction to the media campaigns or to assess changes in knowledge, attitudes, and behavior following instruction. Articulating the logic of their programs helped them identify expected short-term, intermediate, and long-term outcomes and how to measure them. However, only EPA developed an approach for measuring the environmental outcomes of desired behavior changes. Most of the programs we reviewed assumed that program exposure or participation was responsible for observed behavioral changes and failed to address the influence of external factors. The National Youth Anti-Drug Media Campaign evaluation used statistical controls to limit the influence of other factors on its desired outcomes.

Congressional interest was key to initiating most of these evaluations; collaboration with program partners, previous research, and evaluation expertise helped carry them out. Congressional concern about program effectiveness spurred two formal evaluation mandates and other program assessment activities. Collaborations helped ensure that an evaluation would meet the needs of diverse stakeholders. Officials used existing research to design program strategies and establish links to agency goals. Agency evaluation expertise and logic models guided several evaluations in articulating program strategy and expected outcomes. Other agencies could benefit from following the evaluation strategies we describe in this report when they evaluate their information campaigns.

Background

Federal agencies are increasingly expected to demonstrate how their activities contribute to achieving agency or governmentwide goals. The Government Performance and Results Act of 1993 requires federal agencies to report annually on their progress in achieving their agency and program goals. In spring 2002, the Office of Management and Budget (OMB) launched an effort as part of the President's Budget and Performance Integration Management Initiative to highlight what is known about program results. Formal effectiveness ratings for 20 percent of federal programs will initially be conducted under the executive budget formulation process for fiscal year 2004. However, agencies have had difficulty assessing outcomes that are not quickly achieved or readily observed or over which they have little control.

One type of program whose effectiveness is difficult to assess attempts to achieve social or environmental outcomes by informing or persuading others to take actions that are believed to lead to those outcomes. Examples are media campaigns to encourage health-promoting behavior and instruction in adopting practices to reduce environmental pollution. Their effectiveness can be difficult to evaluate because their success depends on the effectiveness of several steps that entail changing knowledge, awareness, and individual behavior that result in changed health conditions or environmental conditions. These programs are expected to achieve their goals in the following ways:

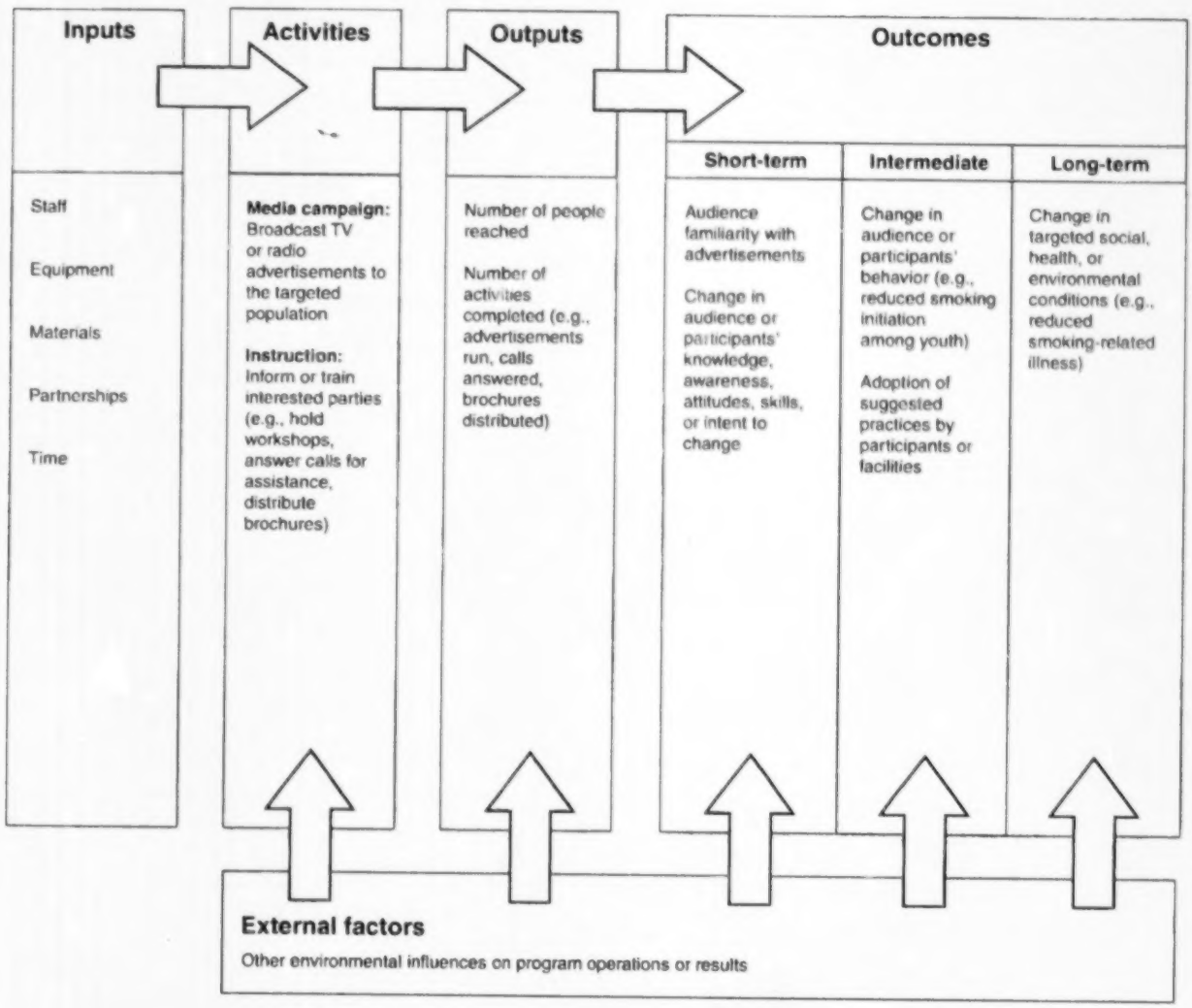
- The program will provide information about a particular problem, why it is important, and how the audience can act to prevent or mitigate it.
- The audience hears the message, gains knowledge, and changes its attitude about the problem and the need to act.

- The audience changes its behavior and adopts more effective or healthful practices.
- The changed behavior leads to improved social, health, or environmental outcomes for the audience individually and, in the aggregate, for the population or system.

How this process can work is viewed from different perspectives. Viewed as persuasive communication, the characteristics of the person who presents the message, the message itself, and the way it is conveyed are expected to influence how the audience responds to and accepts the message. Another perspective sees the targeting of audience beliefs as important factors in motivating change. Still another perspective sees behavior change as a series of steps—increasing awareness, contemplating change, forming an intention to change, actually changing, and maintaining changed behavior. Some programs assume the need for some of but not all these steps and assume that behavior change is not a linear or sequential process. Thus, programs operate differently, reflecting different assumptions about what fosters or impedes the desired outcome or desired behavior change. Some programs, for example, combine information activities with regulatory enforcement or other activities to address factors that are deemed critical to enabling change or reinforcing the program's message.

A program logic model is an evaluation tool used to describe a program's components and desired results and explain the strategy—or logic—by which the program is expected to achieve its goals. By specifying the program's theory of what is expected at each step, a logic model can help evaluators define measures of the program's progress toward its ultimate goals. Figure 1 is a simplified logic model for two types of generic information dissemination programs.

Figure 1: Information Dissemination Program Logic Model


Inputs
Activities
Outputs
Outcomes
Short-term
Intermediate
Long-term
Staff
Equipment
Materials
Partnerships
Time
Media campaign: Broadcast TV or radio advertisements to the targeted population
Instruction: Inform or train interested parties (e.g., hold workshops, answer calls for assistance, distribute brochures)
Number of people reached
Number of activities completed (e.g., advertisements run, calls answered, brochures distributed)
Audience familiarity with advertisements
Change in audience or participants' knowledge, awareness, attitudes, skills, or intent to change
Change in audience or participants' behavior (e.g., reduced smoking initiation among youth)
Adoption of suggested practices by participants or facilities
Change in targeted social, health, or environmental conditions (e.g., reduced smoking-related illness)
External factors
Other environmental influences on program operations or results

Source: GAO's analysis.

BEST COPY AVAILABLE

A program evaluation is a systematic study using objective measures to analyze how well a program is working. An evaluation that examines how a program was implemented and whether it achieved its short-term and intermediate results can provide important information about why a program did or did not succeed on its long-term results. Scientific research methods can help establish a causal connection between program activities and outcomes and can isolate the program's contribution to them. Evaluating the effectiveness of information dissemination programs entails answering several questions about the different stages of the logic model:

- Short-term outcomes: Did the audience consider the message credible and worth considering? Were there changes in audience knowledge, attitudes, and intentions to change behavior?
- Intermediate outcomes: Did the audience's behavior change?[1]
- Long-term outcomes: Did the desired social, health, or environmental conditions come about?

Scope and Methodology

To identify ways that agencies can evaluate how their information dissemination programs contribute to their goals, we conducted case studies of how five agencies evaluate their media campaign or instructional programs. To select the cases, we reviewed departmental and agency performance plans and reports and evaluation reports. We selected cases to represent a variety of evaluation approaches and methods. Four of the cases consisted of individual programs; one represented an office assisting several programs. We describe all five cases in the next section.

To identify the analytic challenges that the agencies faced, we reviewed agency and program materials. We confirmed our understanding with agency officials and obtained additional information on the circumstances that led them to conduct their evaluations. Our findings are limited to the examples reviewed and thus do not necessarily reflect the full scope of these programs' or agencies' evaluation activities.

We conducted our work between October 2001 and July 2002 in accordance with generally accepted government auditing standards.

[1]Some intermediate behavioral outcomes may occur in the short term.

We requested comments on a draft of this report from the heads of the agencies responsible for the five cases. The U.S. Department of Agriculture (USDA), the Department of Health and Human Services (HHS), and EPA provided technical comments that we incorporated where appropriate throughout the report.

Case Descriptions

We describe the goals, major activities, and evaluation approaches and methods for the five cases in this section.

EPA Compliance Assistance

EPA's Compliance Assistance Program disseminates industry-specific and statute-specific information to entities that request it to help them gain compliance with EPA's regulations and thus improve environmental performance. Overseen and implemented by the Office of Enforcement and Compliance Assurance (OECA) and regional offices, compliance assistance consists of telephone help lines, self-audit checklists, written guides, expert systems, workshops, and site visits of regulated industries. OECA provides regional offices with evaluation guidance that illustrates how postsession surveys and administrative data can be used to assess changes in knowledge or awareness of relevant regulations or statutes and adoption of practices. EPA encourages the evaluation of local projects to measure their contribution to achieving the agency's environmental goals.

Eisenhower Professional Development Program

In the U.S. Department of Education, the Eisenhower Professional Development Program supports instructional activities to improve the quality of elementary and secondary school teaching and, ultimately, student learning and achievement. Part of school reform efforts, the program aims to provide primarily mathematics and science teachers with skills and knowledge to help students meet challenging educational standards. Program funds are used nationwide for flexible professional development activities to address local needs related to teaching practices, curriculum, and student learning styles. The national evaluation conducted a national survey of program coordinators and participating teachers to characterize the range of program strategies and the quality of program-assisted activities. The evaluation also collected detailed data at three points in time from all mathematics and science teachers in 10 sites to assess program effects on teachers' knowledge and teaching practices.

Expanded Food and Nutrition Education Program and Other Cooperative Extension Programs	USDA's Cooperative State Research, Education, and Extension Service (CSREES) conducts EFNEP in partnership with the Cooperative Extension System, a network of educators in land grant universities and county offices. EFNEP is an educational program on food safety, food budgeting, and nutrition to assist low-income families acquire knowledge, skills, and changed behavior necessary to develop nutritionally sound diets and improve the total family diet and nutritional well-being. County extension educators train and supervise paraprofessionals and volunteers, who teach the curriculum of about 10 sessions. EFNEP programs across the country measure participants' nutrition-related behavior at program entry and exit on common instruments and report the data to USDA through a common reporting system. In addition, the Cooperative Extension System conducts a variety of other educational programs to improve agriculture and communities and strengthen families. State cooperative extension staff developed and provided evaluation guidance, supported in part by CSREES, to encourage local cooperative extension projects to assess, monitor, and report on performance. Evaluation guidance, including examples of surveys, was provided in seminars and on Web sites to help extension educators evaluate their workshops and their brochures in the full range of topics, such as crop management and food safety.
National Tobacco Control Program	In HHS, the Centers for Disease Control and Prevention (CDC) aims to reduce youths' tobacco use by funding state control programs and encouraging states to use multiple program interventions, working together in a comprehensive approach. CDC supports various efforts, including media campaigns to change youths' attitudes and social norms toward tobacco and to prevent the initiation of smoking. Florida, for example, developed its own counteradvertising, anti-tobacco mass media "truth" campaign. CDC supports the evaluation of local media programs through funding and technical assistance and with state-based and national youth tobacco surveys that provide tobacco use data from representative samples of students. CDC also provides general evaluation guidance for grantee programs to assess advertisement awareness, knowledge, attitudes, and behavior.
National Youth Anti-Drug Media Campaign	The Office of National Drug Control Policy (ONDCP) in the Executive Office of the President oversees the National Youth Anti-Drug Media Campaign, which aims to educate and enable youths to reject illegal drugs. This part of the nation's drug control strategy uses a media campaign to counteract images that are perceived as glamorizing or condoning drug use and to encourage parents to discuss drug abuse with their children.

The media campaign, among other activities, consists of broadcasting paid advertisements and public service announcements that support good parenting practices and discourage drug abuse. While ONDCP oversees the campaign in conjunction with media and drug abuse experts, advertising firms and nonprofit organizations develop the advertisements, which are broadcast to the target audience several times a week for several weeks or months across various media (TV, radio, newspapers, magazines, and billboards) at multiple sites nationwide. The ongoing national evaluation is being conducted by a contractor under the direction of the National Institute on Drug Abuse (NIDA). The evaluation surveys households in the target markets to assess advertisement awareness, knowledge, attitudes, and behavior, including drug use, in a representative sample of youths and their parents or other caretakers.

Program Flexibility, Delayed Effects, and External Influences Posed Major Evaluation Challenges

The programs we reviewed faced challenges to evaluating effects at each step, from conveying information to achieving social and environmental goals. Specifically,

- Flexible programs were hard to summarize nationally as they varied their activities, message, and goals to meet local needs.
- Mass media campaigns do not readily know whether their targeted audience heard the program's message.
- Intended changes in knowledge, attitude, and behavior did not necessarily take place until after audience contact with the program and were, therefore, difficult to observe.
- Self-reports of knowledge, attitudes, and behavior can be prone to bias.
- Long-term behavioral changes and environmental, health, or other social outcomes can take a long time to develop.
- Many factors aside from the program are expected to contribute to the desired behavioral changes and long-term outcomes.

Local Program Variability Makes Nationwide Evaluation Difficult

Several programs we reviewed have broad, general goals and delegated to state or local agencies the authority to determine how to carry out the programs to meet specific local needs. For two reasons, the resulting variability in activities and goals across communities constrained the federal agencies' ability to construct national evaluations of the programs.

First, when states and localities set their own short-term and intermediate goals, common measures to aggregate across projects are often lacking, so it is difficult to assess national progress toward a common goal. Second, these programs also tended to have limited federal reporting requirements. Thus, little information was available on how well a national program was progressing toward national goals.

The Eisenhower Professional Development Program, National Tobacco Control Program, EPA's Compliance Assistance, and CSREES provide financial assistance to states or regional offices with limited federal direction on activities or goals. Many decisions about who receives services and what services they receive are made largely at the regional, county, or school district levels. For example, in the Eisenhower Professional Development Program, districts select professional development activities to support their school reform efforts, including alignment with state and local academic goals and standards. These standards vary, some districts having more challenging standards than others. In addition, training may take various forms; participation in a 2-hour workshop is not comparable to involvement in an intensive study group or year-long course. Such differences in short-term goals, duration, and intensity make counting participating teachers an inadequate way to portray the national program. Such flexibility enables responsiveness to local conditions but reduces the availability of common measures to depict a program in its entirety.

These programs also had limited federal reporting requirements. Cooperative extension and regional EPA offices are asked to report monitoring data on the number of workshops held and clients served, for example, but only selected information on results. The local extension offices are asked to periodically report to state offices monitoring data and accomplishments that support state-defined goals. The state offices, in turn, report to the federal office summary data on their progress in addressing state goals and how they fit into USDA's national goals. The federal program may hold the state and local offices accountable for meeting their state's needs but may have little summary information on progress toward achieving USDA's national goals.

Media Campaigns Lack Interaction with Their Audience

Media campaigns base the selection of message, format, and frequency of broadcast advertisements on audience analysis to obtain access to a desired population. However, a campaign has no direct way of learning whether it has actually reached its intended audience. The mass media campaigns ONDCP and CDC supported had no personal contact with their youth audiences while they received messages from local radio, TV, and

billboard advertisers. ONDCP campaign funds were used to purchase media time and space for advertisements that were expected to deliver two to three anti-drug messages a week using various types of media to the average youth or parent. However, the campaign did not automatically know what portions of the audience heard or paid any attention to the advertisements or, especially, changed their attitudes as a result of the advertisements.

Changes in Behavior Take Place at Home or Work

The instructional programs had the opportunity to interact with their audience and assess their knowledge, skills, and attitudes through questionnaires or observation. However, while knowledge and attitudes may change during a seminar, most desired behavior change is expected to take place when the people attending the seminar return home or to their jobs. Few of these programs had extended contact with their participants to observe such effects directly. In the Eisenhower program, a teacher can learn and report an intention to adopt a new teaching practice, but this does not ensure that the teacher will actually use it in class.

Participants' Self-Reports May Produce Poor-Quality Data

End-of-session surveys asking for self-reports of participants' knowledge, attitudes, and intended behavior are fast and convenient ways to gain information but can produce data of poor quality. This can lead to a false assessment of a workshop's impact. Respondents may not be willing to admit to others that they engage in socially sensitive or stigmatizing activities like smoking or drug use. They may not trust that their responses will be kept confidential. In addition, they may choose to give what they believe to be socially desirable or acceptable answers in order to appear to be doing the "right thing." When surveys ask how participants will use their learning, participants may feel pressured to give a positive but not necessarily truthful report. Participants may also report that they "understand" the workshop information and its message but may not be qualified to judge their own level of knowledge.

Outcomes Take Time to Develop

Assessing a program's intermediate behavioral outcomes, such as smoking, or long-term outcomes, such as improved health status, is hindered by the time they take to develop. To evaluate efforts to prevent youths from starting to smoke, evaluators need to wait several years to observe evidence of the expected outcome. ONDCP expects its media campaign to take about 2 to 3 years to affect drug use. Many population-based health effects take years to become apparent, far beyond the reach of these programs to study.

Tracking participants over several years can be difficult and costly. Even after making special efforts to locate people who have moved, each year a few more people from the original sample may not be reached or may refuse to cooperate. In the Eisenhower evaluation, 50 percent of the initial sample (60 percent of teachers remaining in the schools) responded to all three surveys. When a sample is tracked for several years, the cumulative loss of respondents may eventually yield such a small proportion of the original sample as not to accurately represent that original sample. Moreover, the proportion affected tends to diminish at each step of the program logic model, which can reduce the size of the expected effect on long-term outcomes so small as to be undetectable. That is, if the program reached half the targeted audience, changed attitudes among half of those it reached, half of those people changed their behavior, and half of those experienced improved health outcomes, then only one-sixteenth of the initial target audience would be expected to experience the desired health outcome. Thus, programs may be unlikely to invest in tracking the very large samples required to detect an effect on their ultimate outcome.

Other Factors Influence Desired Outcomes

Attributing observed changes in participants to the effect of a program requires ruling out other plausible explanations. Those who volunteer to attend a workshop are likely to be more interested, knowledgeable, or willing to change their behavior than others who do not volunteer. Environmental factors such as trends in community attitudes toward smoking could explain changes in youths' smoking rates. ONDCP planners have recognized that sensation seeking among youths is associated with willingness to take social or physical risks; high-sensation seekers are more likely to be early users of illegal drugs. Program participants' maturing could also explain reductions in risky behavior over time.

Other programs funded with private or other federal money may also strive for similar goals, making it difficult to separate out the information program's unique contribution. The American Legacy Foundation, established by the 1998 tobacco settlement, conducted a national media campaign to discourage youths from smoking while Florida was carrying out its "truth" campaign. Similarly, the Eisenhower program is just one of many funding sources for teacher development, but it is the federal government's largest investment solely in developing the knowledge and skills of classroom teachers. The National Science Foundation also funds professional development initiatives in mathematics and science. The evaluation found that local grantees combine Eisenhower grants with other funds to pay for conferences and workshops.

Surveys and Logic Models Helped Address Most Challenges, but External Factors Were Rarely Addressed

The agencies we reviewed used a variety of strategies to address their evaluation challenges. Two flexible programs developed common, national measures, while two others promoted locally tailored evaluations. Most programs used exit or follow-up surveys to gather data on short-term and intermediate outcomes. Articulating a logic model for their programs helped some identify appropriate measures and strategies to address their challenges. Only EPA developed an approach for measuring its program's long-term health and environmental outcomes or benefits. Most of the programs we reviewed assumed that program exposure or participation was responsible for observed changes and failed to address the role of external factors. However, the NIDA evaluation did use evaluation techniques to limit the influence of nonprogram factors. Table 1 displays the strategies the five cases used or recommended in guidance to address the challenges.

Table 1: The Programs' Challenges and Their Strategies

Challenge	Strategy
Flexible programs were hard to summarize nationally as they varied their activities, messages, and goals to meet local needs.	• Develop common measures for national program evaluation. • Encourage local projects to evaluate progress toward their own goals.
Mass media campaigns do not readily know whether their target audience heard the program's message.	• Survey intended audience to ask about program exposure, knowledge and attitude change.
Intended changes in knowledge, attitude, and behavior might not take place until after contact with the program and were thus difficult to observe.	• Conduct postworkshop survey or follow-up surveys. • Conduct observations. • Use existing administrative or site visit data.
Self-report surveys of knowledge, attitudes, or behavior can be prone to bias.	• Adjust wording of survey questions. • Ensure confidentiality of survey and its results. • Compare before-and-after reports to assess change.
Long-term behavioral changes and environmental, health, or other social outcomes can take a long time to develop.	• Assess intermediate outcomes. • Use logic model to demonstrate links to agency goals. • Conduct follow-up survey.
Many factors aside from the program are expected to contribute to the desired behavioral changes and long-term outcomes.	• Select outcomes closely associated with the program. • Use statistical methods to limit external influences. • Evaluate the combined effect of related activities rather than trying to limit their influences.

Source: GAO's analysis.

Find Common Measures or Encourage Locally Tailored Evaluations

Two of the four flexible programs developed ways to assess progress toward national program goals, while the others encouraged local programs to conduct their own evaluations, tailored to local program goals.

EFNEP does not have a standard national curriculum, but local programs share common activities aimed at the same broad goals. A national committee of EFNEP educators developed a behavior checklist and food

recall log to provide common measures of client knowledge and adoption of improved nutrition-related practices, which state and local offices may choose to adopt. The national program office provided state and local offices with software to record and analyze client data on these measures and produce tailored federal and state reports. In contrast, lacking standard reporting on program activities or client outcomes, the Eisenhower program had to conduct a special evaluation study to obtain such data. The evaluation contractor surveyed the state program coordinators to learn what types of training activities teachers were enrolled in and surveyed teachers to learn about their training experiences and practices. The evaluation contractor drew on characteristics identified with high-quality instruction in the research literature to define measures of quality for this study.

In contrast, EPA and CDC developed guidance on how to plan and conduct program evaluations and encouraged state and local offices to assess their own individual efforts. To measure the effects of EPA's enforcement and compliance assurance activities, the agency developed a performance profile of 11 sets of performance measures to assess the activities undertaken (including inspections and enforcement, as well as compliance assistance), changes in the behavior of regulated entities, and progress toward achieving environmental and health objectives. One set of measures targets the environmental or health effects of compliance assistance that must be further specified to apply to the type of assistance and relevant industry or sector. However, EPA notes that since the measured outcomes are very specific to the assistance tool or initiative, aggregating them nationally will be difficult. Instead, EPA encourages reporting the outcomes as a set of quantitative or qualitative accomplishments.

In CDC's National Tobacco Control Program, states may choose to conduct any of a variety of activities, such as health promotions, clinical management of nicotine addiction, advice and counseling, or enforcing regulations limiting the access minors have to tobacco. With such intentional flexibility and diversity, it is often difficult to characterize or summarize the effectiveness of the national program. Instead, CDC conducted national and multistate surveillance, providing both baseline and trend data on youths' tobacco use, and encouraged states to evaluate their own programs, including surveying the target audience's awareness and reactions. CDC's "how to" guide assists program managers and staff in planning and implementing evaluation by providing general evaluation guidance that includes example outcomes—short term, intermediate, and

long term—and data sources for various program activities or interventions.[2]

Survey the Population Targeted by the Media Campaign

Both mass media campaigns surveyed their intended audience to learn how many heard or responded to the message and, thus, whether the first step of the program was successful. Such surveys, a common data source for media campaigns, involved carefully identifying the intended audience, selecting the survey sample, and developing the questionnaire to assess the intended effects.

The National Youth Anti-Drug Media Campaign is designed to discourage youths from beginning to use drugs by posting advertisements that aim to change their attitudes about drugs and encourage parents to help prevent their children from using drugs. Thus, the NIDA evaluation developed a special survey, the National Survey of Parents and Youth (NSPY), with parallel forms to address questions about program exposure and effects on both groups. At the time of our interview, NSPY had fielded three waves of interviews to assess initial and cumulative responses to the campaign but planned additional follow-up. Cross-sectional samples of youths and parents (or caregivers) were drawn to be nationally representative and produce equal-sized samples within three age subgroups of particular interest (youths aged 9–11, 12–13, and 14–18). Separate questionnaires for youths and parents measured their exposure to both specific advertisements and, more generally, the campaign and other noncampaign anti-drug messages. In addition, they were asked about their beliefs, attitudes, and behavior regarding drug use and factors known to be related to drug use (for youths) or their interactions with their children (for parents).

Florida's tobacco control program integrated an advertisement campaign to counter the tobacco industry's marketing with community involvement, education, and enforcement activities. The campaign disseminates its message about tobacco industry advertising through billboards and broadcasting and by distributing print media and consumer products (such as hats and T-shirts) at events for teenagers. Florida's Anti-tobacco Media Evaluation surveys have been conducted every 6 months since the

[2]Goldie MacDonald and others, *Introduction to Program Evaluation for Comprehensive Tobacco Control Programs* (Atlanta, Ga.: Centers for Disease Control and Prevention, November 2001).

program's inception in 1998 to track awareness of the campaign as well as youths' anti-tobacco attitudes, beliefs, and smoking behavior.

Assess Postworkshop Changes with Surveys and Observations

Most of the instructional programs we reviewed assessed participants' short-term changes in knowledge, attitudes, or skills at the end of their session and relied on follow-up surveys to learn about intermediate effects that took place later. EFNEP and EPA's Compliance Assistance, which had more extended contact with participants, were able to collect more direct information on intermediate behavioral effects.

State cooperative extension and EPA evaluation guidance encouraged program staff to get immediate feedback on educational workshops, seminars, and hands-on demonstrations and their results. Reference materials suggested that postworkshop surveys ask what people think they gained or intend to do as a result of the program sessions.[3] Questions may ask about benefits in general or perceived changes in specific knowledge, skills, attitudes, or intended actions. These surveys can show postprogram changes in knowledge and attitudes but not whether the participants actually changed their behavior or adopted the recommended practices. An extension evaluator said that this is the typical source of evaluation data for some types of extension programs.

Cooperative extension evaluations have also used other types of on-site data collection, such as observation during workshops to document how well participants understood and can use what was taught.[4] The traditional paper-and-pencil survey may be less effective with children or other audiences with little literacy, so other sources of data are needed. Program or evaluation staff can observe (directly or from documents) the use of skills learned in a workshop—for example, a mother's explaining to another nonparticipating mother about the need to wash hands before food preparation. Staff can ask participants to role-play a scenario—for example, an 8-year-old's saying "no" to a cigarette offered by a friend. These observations could provide evidence of knowledge, understanding

[3]See, for example, Ellen Taylor-Powell and Marcus Renner, "Collecting Evaluation Data: End-of-Session Questionnaires," University of Wisconsin Cooperative Extension document G3658-11, Madison, Wisconsin, September 2000. Also see the Bibliography for various sources of guidance.

[4]See, for example, Ellen Taylor-Powell and Sara Steele, "Collecting Evaluation Data: Direct Observation," University of Wisconsin Cooperative Extension document G3658-5, Madison, Wisconsin, 1996.

of the skills taught, and ability to act on the message.[5] While these data may be considered more accurate indicators of knowledge and skill gains than self-report surveys, they are more resource-intensive to collect and analyze.

Most of the programs we reviewed expected the desired behavior change—the intermediate outcome—to take place later, after participants returned home or to their jobs. EFNEP is unusual in using surveys to measure behavior change at the end of the program. This is possible because (1) the program collects detailed information on diet, budgeting, and food handling from participants at the start and end of the program and (2) its series of 10 to 12 lessons is long enough to expect to see such changes.

Programs that did not expect behavior to change until later or at work used follow-up surveys to identify actual change in behavior or the adoption of suggested practices. Cooperative extension and EPA's Compliance Assistance evaluation guidance encouraged local evaluators to send a survey several weeks or months later, when participants are likely to have made behavior changes. Surveys may be conducted by mail, telephone, or online, depending on what appears to be the best way to reach potential respondents. An online survey of Web site visitors, for example, can potentially reach a larger number of respondents than may be known to the program or evaluator. EPA recommended that the form of evaluation follow-up match the form and intensity of the intervention, such as conducting a periodic survey of a sample of those who seek assistance of a telephone help-desk rather than following up each contact with an extensive survey. EPA and ONDCP officials noted that survey planning must accommodate a review by the Office of Management and Budget to ascertain whether agency proposals for collecting information comply with the Paperwork Reduction Act.[6]

EPA guidance encouraged evaluators to obtain administrative data on desired behavior changes rather than depending on less-reliable self-report survey data. Evidence of compliance can come from observations during follow-up visits to facilities that had received on-site compliance assistance or from tracking data that the audience may be required to

[5]Nancy Ellen Kiernan, "Using Observation to Evaluate Skills," Penn State University Cooperative Extension Tipsheet 61, University Park, Pennsylvania, 2001.

[6]44 U.S.C. 3501-3520 (2000).

report for regulatory enforcement purposes. For example, after a workshop for dry cleaners about the permits needed to meet air quality regulations, EPA could examine data on how many of the attendees applied for such permits within 6 months after the workshop. This administrative data could be combined with survey results to obtain responses from many respondents yet collect detailed information from selected participants.

Adjust Self-Report Surveys to Reduce Potential Bias

Using a survey at the end of a program session to gain information from a large number of people is fast and convenient, but self-reports may provide positively biased responses about the session or socially sensitive or controversial topics. To counteract these tendencies, the programs we reviewed used various techniques either to avoid threatening questions that might elicit a socially desirable but inaccurate response or to reassure interviewees of the confidentiality of their responses. In addition, the programs recommended caution in using self-reports of knowledge or behavior changes, encouraging evaluators—rather than participants—to assess change.

Carefully wording questions can encourage participants to candidly record unpopular or negative views and can lessen the likelihood of their giving socially desirable responses. Cooperative extension evaluation guidance materials suggest that survey questions ask for both program strengths and weaknesses or for suggestions on how to improve the program. These materials also encourage avoidance of value-laden terms. Questions about potentially embarrassing situations might be preceded by a statement that acknowledges that this happens to everyone at some time.[7]

To reassure respondents, agencies also used the survey setting and administration to provide greater privacy in answering the questions. Evaluation guidance encourages collecting unsigned evaluation forms in a box at the end of the program, unless, of course, individual follow-up is desired. Because the National Youth Anti-Drug Media Campaign was dealing with much more sensitive issues than most surveys, its evaluation took several steps to reassure respondents and improve the quality of the data it collected. Agency officials noted that decisions about survey design and collecting quality data involve numerous issues such as consent,

[7]For a review of related research see Norbert Schwarz and Daphna Oyserman, "Asking Questions about Behavior: Cognition, Communication, and Questionnaire Construction," *American Journal of Evaluation* 22:2 (summer 2001): 127–60.

parental presence, feasibility, mode, and data editing procedures. In this case, they chose a panel study with linked data from youths and one parent or guardian collected over three administrations. In addition, they found that obtaining cooperation from a representative sample of schools with the frequency required by the evaluation was not feasible. So the evaluation team chose to survey households in person instead of interviewing youths at school or conducting a telephone survey.

Hoping to improve the quality of sensitive responses, the surveyors promised confidentiality and provided respondents with a certificate of confidentiality from HHS. In addition, the sensitive questions were self-administered with a touch-screen laptop computer. All sensitive questions and answer categories appeared on the laptop screen and were spoken to the respondent by a recorded voice through earphones. Respondents chose responses by touching the laptop screen. This audio computer-assisted self-interview instrument was likely to obtain more honest answers about drug use, because respondents entered their reports without their answers being observed by the interviewer or their parents. NIDA reported that a review of the research literature on surveys indicated that this method resulted in higher reported rates of substance abuse for youths, compared to paper-and-pencil administration.

Compare Presession and Postsession Reports to Assess Change

State cooperative extension and EPA evaluation guidance cautioned that self-reports may not reflect actual learning or change; they encouraged local projects to directly test and compare participant knowledge before and after an activity rather than asking respondents to report their own changed behavior. Both the EFNEP and Eisenhower evaluators attempted to reduce social desirability bias in self-reports of change by asking for concrete, detailed descriptions of what the respondents did before and after the program. By asking for a detailed log of what participants ate the day before, EFNEP sought to obtain relatively objective information to compare with nutrition guidelines. By repeating this exercise at the beginning and end of the program, EFNEP obtained more credible evidence than by asking participants whether they had adopted desired practices, such as eating less fat and more fruit and vegetables.

The Eisenhower evaluation also relied on asking about very specific behaviors to minimize subjectivity and potential bias. First, evaluators analyzed detailed descriptions of their professional development activities along characteristics identified as important to quality in prior research—such as length and level of involvement. Thus, they avoided asking teachers to judge the quality of their professional development activities. Second, teachers were surveyed at three points in time to obtain detailed

information on their instructional practices during three successive school years. Teachers were asked to complete extensive tables on the content and pedagogy used in their course; then the evaluators analyzed whether these represented high standards and effective instructional approaches as identified in the research literature. The evaluators then compared teacher-reported instructional practices before and after their professional development training to assess change on key dimensions of quality.

Some cooperative extension guidance noted that pretest-posttest comparison of self-report results may not always provide accurate assessment of program effects, because participants may have limited knowledge at the beginning of the program that prevents them from accurately assessing baseline behaviors. For example, before instruction on the sources of certain vitamins, participants may inaccurately assess the adequacy of their own consumption levels. The "post-then-pre" design can address this problem by asking participants to report at the end of the program, when they know more about their behavior, both then and as it was before the program. Evidently, participants may also be more willing to admit to certain inappropriate behaviors.[8]

Use Program Logic Models to Show Links to Unmeasured Long-Term Outcomes

Assessing long-term social or health outcomes that were expected to take more than 2 to 3 years to develop was beyond the scope of most of these programs. Only EPA developed an approach for measuring long-term outcomes, such as the environmental effects of desired behavior change in cases where they can be seen relatively quickly. In most instances, programs measured only short-term and intermediate outcomes, which they claimed would contribute to achieving these ultimate benefits. Several programs used logic models to demonstrate their case; some drew on associations established in previous research. The Eisenhower and NIDA evaluations took special effort to track participants long enough to observe desired intermediate outcomes.

EFNEP routinely measures intermediate behavioral outcomes of improved nutritional intake but does not regularly assess long-term outcomes of nutritional or health status, in part because they can take many years to develop. Instead, the program relies on the associations established in

[8]Nancy Ellen Kiernan, "Reduce Bias with Retrospective Questions," Penn State Cooperative Extension Tipsheet 30, University Park, Pennsylvania, 2001, and S. Kay Rockwell and Harriet Kohn, "Post-Then-Pre Evaluation," *Journal of Extension* 27:2 (summer 1989).

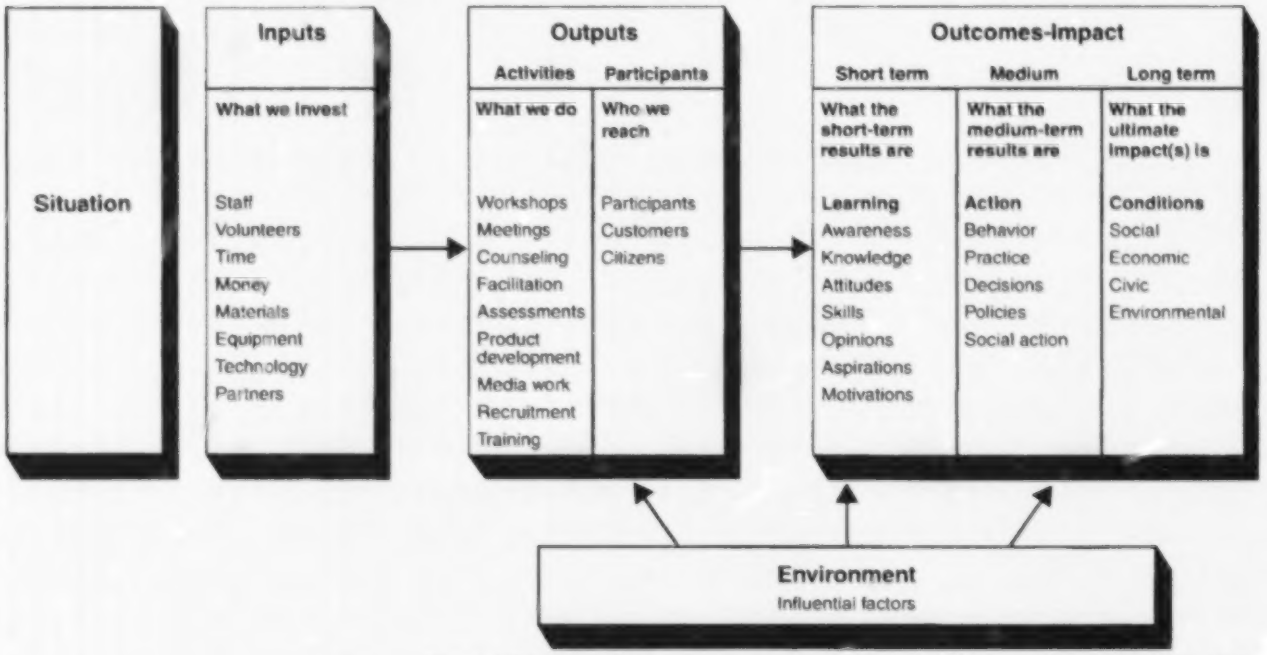
medical research between diet and heart disease and certain cancers, for example, to explain how it expects to contribute to achieving disease-reduction goals. Specifically, Virginia Polytechnic Institute and State University (Virginia Tech) and Virginia cooperative extension staff developed a model to conduct a cost-benefit analysis of the health-promoting benefits of its EFNEP program. The study used equations estimating the health benefits of the program's advocated nutritional changes for each of 10 nutrition-related diseases (such as colorectal cancer) from medical consensus reports. The study then used program data on the number of participants who adopted the whole set of targeted behaviors to calculate the expected level of benefits, assuming they maintained the behaviors for 5 years.

EPA provided regional staff with guidance that allows them to estimate environmental benefits from pollution reduction in specific cases of improved compliance with EPA's regulations. To capture and document the environmental results and benefits of concluded enforcement cases, EPA developed a form for regional offices to record their actions taken and pollutant reductions achieved. The guidance provides steps, formulas, and look-up tables for calculating pollutant reduction or elimination for specific industries and types of water, air, or solid waste regulations.[9] EPA regional staff are to measure average concentrations of pollutants before a specific site becomes compliant and to calculate the estimated total pollutant reduction in the first year of postaction compliance. Where specific pollution-reduction measures can be aggregated across sites, EPA can measure effects nationally and show the contribution to agencywide pollution-reduction goals. In part because these effects occur in the short term, EPA was unique among our cases in having developed an approach for measuring the effects of behavior change.

[9]EPA, Office of Enforcement and Compliance Assurance, *Case Conclusion Data Sheet*, document 2222A (Washington, D.C.: November 2000).

Logic models helped cooperative extension programs and the evaluation of ONDCP's media campaign identify their potential long-term effects and the route through which they would be achieved. The University of Wisconsin Cooperative Extension guidance encourages the use of logic models to link investments to results. They aim to help projects clarify linkages among program components; focus on short-term, intermediate, and long-term outcomes; and plan appropriate data collection and analysis. The guidance suggests measuring outcomes over which the program has a fair amount of control and considering, for any important long-term outcome, whether it will be attained if the other outcomes are achieved. Figure 2 depicts a generic logic model for an extension project, showing how it can be linked to long-term social or environmental goals.

Figure 2: University of Wisconsin Cooperative Extension Logic Model

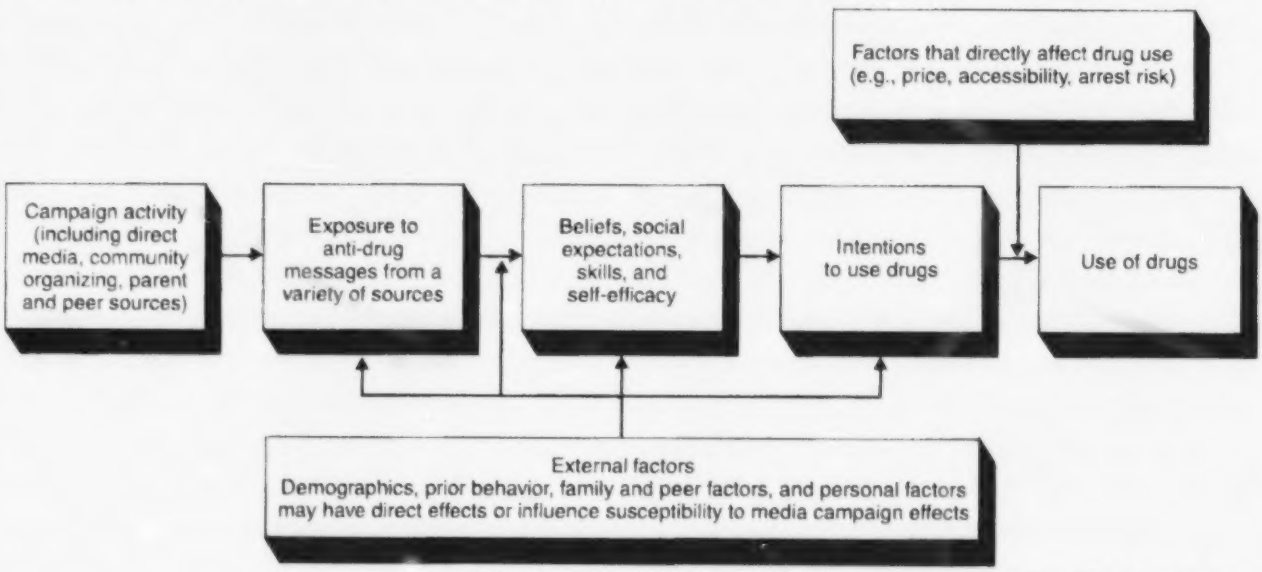


Source: Adapted from Ellen Taylor-Powell, "The Logic Model: A Program Performance Framework," University of Wisconsin Cooperative Extension, Madison, Wisconsin, n.d., http://www.uwex.edu/ces/pdande (September 2002).

BEST COPY AVAILABLE

The evaluation of the National Youth Anti-Drug Media Campaign followed closely the logic of how the program was expected to achieve its desired outcomes, and its logic models show how the campaign contributes to ONDCP's drug-use reduction goals. For example, the campaign had specific hypotheses about the multiple steps through which exposure to the media campaign message would influence attitudes and beliefs, which would then influence behavior. Thus, evaluation surveys tapped various elements of youths' attitudes and beliefs about drug use and social norms, as well as behaviors that are hypothesized to be influenced by—or to mediate the influence of—the campaign's message. In addition, NIDA plans to follow for 2 to 3 years those who had been exposed to the campaign to learn how the campaign affected their later behavior. Figure 3 shows the multiple steps in the media campaign's expected influence and how personal factors affect the process.

Figure 3: Logic Model for the National Youth Anti-Drug Media Campaign Evaluation


Factors that directly affect drug use (e.g., price, accessibility, arrest risk)
Campaign activity (including direct media, community organizing, parent and peer sources)
Exposure to anti-drug messages from a variety of sources
Beliefs, social expectations, skills, and self-efficacy
Intentions to use drugs
Use of drugs
External factors
Demographics, prior behavior, family and peer factors, and personal factors may have direct effects or influence susceptibility to media campaign effects

Source: Adapted from Robert Hornik and others, Evaluation of the National Youth Anti-Drug Media Campaign: Historical Trends in Drug Use and Design of the Phase III Evaluation, prepared for the National Institute on Drug Abuse (Rockville, Md.: Westat, July 2000).

BEST COPY AVAILABLE

Following program participants for years to learn about the effects on long-term outcomes for specific individuals exceeded the scope of most of these programs; only the formal evaluation studies of the Eisenhower and ONDCP programs did this. It can be quite costly to repeatedly survey a group of people or track individuals' locations over time and may require several attempts in order to obtain an interview or completed survey. The Eisenhower evaluation employed a couple of techniques that helped reduce survey costs. First, the evaluation increased the time period covered by the surveys by surveying teachers twice in one year: first about their teaching during the previous school year and then about activities in the current school year. By surveying teachers in the following spring about that school year, the evaluators were able to learn about three school years in the space of 1-1/2 actual years. Second, the case study design helped reduce survey costs by limiting the number of locations the evaluation team had to revisit. Concentrating their tracking efforts in 10 sites also allowed the team to increase the sample of teachers and, thus, be more likely to detect small effects on teaching behavior.

Control for External Influences or Assess Their Combined Effects

Most of the evaluations we reviewed assumed that program exposure or participation led to the observed behavioral changes and did not attempt to control the influence of external factors. However, in order to make credible claims that these programs were responsible for a change in behavior, the evaluation design had to go beyond associating program exposure with outcomes to rule out the influence of other explanations. NIDA's evaluation used statistical controls and other techniques to limit the influence of other factors on attitudes and behaviors, while Eisenhower, CDC, and EPA encouraged assessment of the combined effect of related activities aimed at achieving the same goals.

EFNEP's evaluation approach paired program exposure with before-and-after program measures of outcomes to show a change that was presumed to stem from the program. Where the recommended behavior is very specific and exclusive to a program, it can be argued that the program was probably responsible for its adoption. An EFNEP program official explained that because program staff work closely with participants to address factors that could impede progress, they are comfortable using the data to assess their effectiveness.

Many factors outside ONDCP's media campaign were expected to influence youths' drug use, such as other anti-drug programs and youths' willingness to take risks, parental attitudes and behavior, peer attitudes and behavior, and the availability of and access to drugs. NIDA's evaluation used several approaches to limit the effects of other factors on

the behavioral outcomes it was reporting. First, to distinguish this campaign from other anti-drug messages in the environment, the campaign used a distinctive message to create a "brand" that would provide a recognizable element across advertisements in the campaign and improve recall of the campaign. The evaluation's survey asked questions about recognition of this brand, attitudes, and drug use so the analysis could correlate attitudes and behavior changes with exposure to this particular campaign.

Second, NIDA's evaluation used statistical methods to help limit the influence of other factors on the results. The evaluation lacked a control group that was not exposed, since the campaign ran nationally, or baseline data on the audience's attitudes before the campaign began, with which to compare the survey sample's reaction. Thus, the evaluation chose to compare responses to variation in exposure to the campaign—comparing those with high exposure to those with low exposure—to assess its effects. This is called a dose-response design which assesses how risk of disease increases with increasing doses or exposure. This approach presumes that the advertisements were effective if you were more likely to adopt the promoted attitudes or behaviors as you saw more of them.

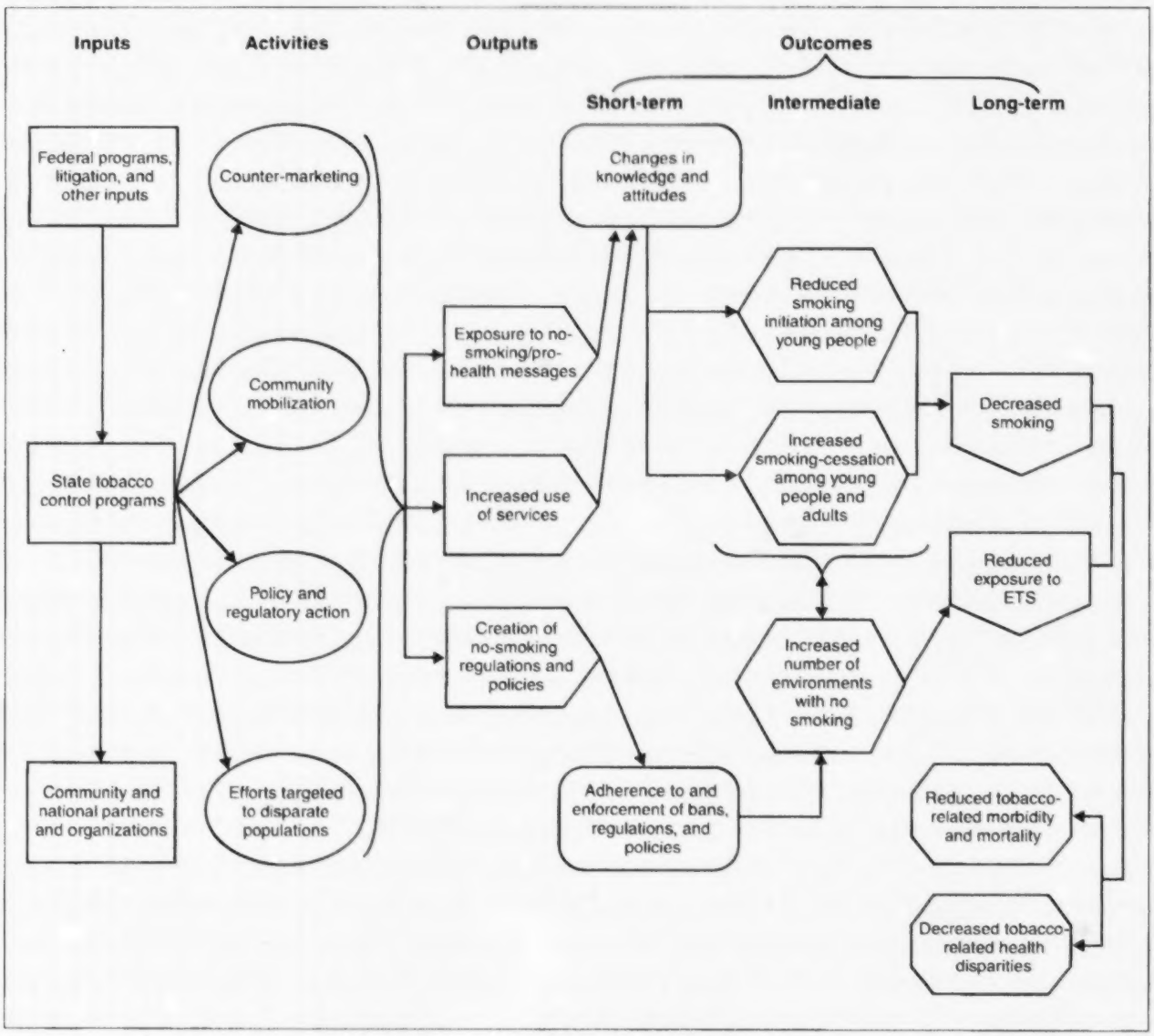
However, because the audience rather than the evaluator determined how many advertisements they saw, it is not a random selection process, and other factors related to drug use may have influenced both audience viewing habits and drug-related attitudes and behaviors. To limit the influence of preexisting differences among the exposure groups on the results, the NIDA evaluation controlled for their influence by using a special statistical method called propensity scoring. This controls for any correlation between program exposure and risk factors for drug use, such as gender, ethnicity, strength of religious feelings, and parental substance abuse, as well as school attendance and participation in sensation-seeking activities. This statistical technique requires detailed data on large numbers of participants and sophisticated analysis resources.

Some information campaigns are intertwined or closely associated with another program or activity aimed at the same goals. Both Eisenhower and the other programs fund teachers' professional development activities that vary in quality, yet they found no significant difference in quality by funding source in their sample. So the evaluation focused instead on assessing the effect of high-intensity activities—regardless of funding source—on teaching practice. EPA's Compliance Assistance program, for example, helps regulated entities comply with regulations along with its regulatory enforcement responsibilities—a factor not lost on the entities

that are regulated. EPA's dual role raises the question of whether any observed improvements in compliance result from assistance efforts or the implied threat of inspections and sanctions. EPA measures the success of its compliance assistance efforts together with those of incentives that encourage voluntary correction of violations to promote compliance and reductions in pollution.

An alternative evaluation approach acknowledged the importance of combining information dissemination with other activities to the total program design and assessed the outcomes of the combined activities. This approach, exemplified by CDC and the public health community, encourages programs to adopt a comprehensive set of reinforcing media and regulatory and other community-based activities to produce a more powerful approach to achieving difficult behavior change. The proposed evaluations seek not to limit the influence of these other factors but to assess their combined effects on reducing tobacco use. CDC's National Tobacco Control Program uses such a comprehensive approach to obtain synergistic effects, making moot the issue of the unique contribution of any one program activity. Figure 4 depicts the model CDC provided to help articulate the combined, reinforcing effects of media and other community-based efforts on reducing tobacco use.

Figure 4: CDC Tobacco Use Prevention and Control Logic Model



Note: ETS = environmental tobacco smoke.

Source: Goldie MacDonald and others. *Introduction to Program Evaluation for Comprehensive Tobacco Control Programs* (Atlanta, Ga.: CDC, November 2001).

Congressional Interest, Collaboration, Available Information and Expertise Supported These Evaluations

Agencies initiated most of these evaluation efforts in response to congressional interest and questions about program results. Then, collaboration with program partners and access to research results and evaluation expertise helped them carry out and increase the contributions of these evaluations.

Congressional Interest

Congressional concern about program effectiveness resulted in two mandated evaluations and spurred agency performance assessment efforts in two others. The Congress encouraged school-based education reform to help students meet challenging academic standards with the Improving America's Schools Act of 1994.[10] Concerned about the quality of professional development to update teaching practices needed to carry out those reforms, the Congress instituted a number of far-reaching changes and mandated an evaluation for the Eisenhower Professional Development Program. The formal 3-year evaluation sought to determine whether and how Eisenhower-supported activities, which constitute the largest federal effort dedicated to supporting educator professional development, contribute to national efforts to improve schools and help achieve agency goals.

The Congress has also been actively involved in the development and oversight of the National Youth Anti-Drug Media Campaign. It specified the program effort in response to nationwide rises in rates of youths' drug use and mandated an evaluation of that effort. ONDCP was asked to develop a detailed implementation plan and a system to measure outcomes of success and report to the Congress within 2 years on the effectiveness of the campaign, based on those measurable outcomes. ONDCP contracted for an evaluation through NIDA to ensure that the evaluation used the best research design and was seen as independent of the sponsoring agency. ONDCP requested reports every 6 months on program effectiveness and impact. However, officials noted that this reporting schedule created unrealistically high congressional expectations

[10]P.L. 103-382, Oct. 20, 1994, 108 Stat. 3518.

for seeing results when the program does not expect to see much change in 6 months.

Congressional interest in sharpening the focus of cooperative extension activities led to installing national goals that were to focus the work and encourage the development of performance goals. The Agricultural Research, Extension, and Education Reform Act of 1998 gave states authority to set priorities and required them to solicit input from various stakeholders.[11] The act also encouraged USDA to address high-priority concerns with national or multistate significance. Under the act, states are required to develop plans of work that define outcome goals and describe how they will meet them. Annual performance reports are to describe whether states met their goals and to report their most significant accomplishments. CSREES draws on these reports of state outcomes to describe how they help meet USDA's goals. State extension officials noted that the Government Performance and Results Act of 1993, as well as increased accountability pressures from their stakeholders, created a demand for evaluations.

EFNEP's performance reporting system was also initiated in response to congressional interest and is used to satisfy this latter act's requirements. USDA staff noted that the House Committee on Agriculture asked for data in 1989 to demonstrate the impact of the program to justify the funding level. On the basis of questions from congressional staff, program officials and extension partners formed a national committee that examined the kinds of information that had already been gathered to respond to stakeholders and developed standard measures of desired client improvements. State reports are tailored to meet their information needs, while CSREES uses the core set of common behavioral items to provide accomplishments for USDA's annual performance report.

Collaboration with Program Partners

In several evaluations we reviewed, collaboration was reported as important for meeting the information needs of diverse audiences and expanding the usefulness of the evaluation. ONDCP's National Youth Anti-Drug Media Campaign was implemented in collaboration with the Partnership for a Drug-Free America and a wide array of nonprofit, public, and private organizations to reinforce its message across multiple outlets. The National Institute on Drug Abuse, with input from ONDCP, designed

[11]P.L. 105-185, June 23, 1998, 112 Stat. 523.

the evaluation of the campaign and drew on an expert panel of advisers in drug abuse prevention and media studies. The evaluation was carried out by a partnership between Westat—bringing survey and program evaluation expertise—and the University of Pennsylvania's Annenberg School for Communication—bringing expertise in media studies. Agency officials noted that through frequent communication with those developing the advertisements and purchasing media time, evaluators could keep the surveys up to date with the most recent airings and provide useful feedback on audience reaction.

The Evaluation/Reporting System represented a collaborative effort among the federal and state programs to demonstrate EFNEP's benefits. USDA staff noted that in the early 1990s, in response to congressional inquiries about EFNEP's effectiveness, a national committee was formed to develop a national reporting system for data on program results. The committee held an expert panel with various USDA nutrition policy experts, arranged for focus groups, and involved state and county EFNEP representatives and others from across the country. The committee started by identifying the kinds of information the states had already gathered to respond to state and local stakeholders' needs and then identified other questions to be answered. The committee developed and tested the behavior checklist and dietary analysis methodology from previous nutrition measurement efforts. The partnership among state programs continues through an annual CSREES Call for Questions that solicits suggestions from states that other states may choose to adopt. USDA staff noted that local managers helped design measures that met their needs, ensuring full cooperation in data collection and the use of evaluation results.

State extension evaluator staff emphasized that collaborations and partnerships were an important part of their other extension programs and evaluations. At one level, extension staff partner with state and local stakeholders—the state natural resource department, courts, social service agencies, schools, and agricultural producers—as programs are developed and implemented. This influences whether and how the programs are evaluated—what questions are asked and what data are collected—as those who helped define the program and its goals have a stake in how to evaluate it. State extension evaluator staff also counted their relationships with their peers in other states as key partnerships that provided peer support and technical assistance. In addition to informal contacts, some staff were involved in formal multi-state initiatives, and many participate in a formal shared interest group of the American Evaluation Association. While we were writing our report, the

association's Extension Education Evaluation Topical Interest Group had more than 160 members, a Web site, and a listserv and held regular meetings (see http://www.danr.ucop.edu/eee-aea/).

Findings from Previous Research

Using research helped agencies develop measures of program goals and establish links between program activities and short-term goals and between short-term and long-term goals. The Eisenhower evaluation team synthesized existing research on teacher instruction to develop innovative measures of the quality of teachers' professional development activities, as well as the characteristics of teaching strategies designed to encourage students' high-order thinking. EFNEP drew on nutrition research to develop standard measures for routine assessment and performance reporting. Virginia Tech's cooperative extension program also drew on research on health care expenses and known risk factors for nutrition-related diseases to estimate the benefits of nutrition education on reducing the incidence and treatment costs of those diseases.

Both the design of ONDCP's National Anti-Drug Media Campaign and its evaluation drew on lessons learned in earlier research. The message and structure of the media campaign were based on a review of research evidence on the factors affecting youths' drug use, effective drug-use prevention practices, and effective public health media campaigns. Agency officials indicated that the evaluation was strongly influenced by the "theory of reasoned action" perspective to explain behavioral change. This perspective assumes that intention is an important factor in determining behavior and that intentions are influenced by attitudes and beliefs. Exposure to the anti-drug messages is thus expected to change attitudes, intentions, and ultimately behavior. Similarly, CDC officials indicated that they learned a great deal about conducting and evaluating health promotion programs from their experience with HIV-AIDS prevention demonstration programs conducted in the late 1980s and early 1990s. In particular, earlier research on health promotions shaped their belief in the increased effectiveness of programs that combine media campaigns with other activities having the same goal.

Evaluation Expertise and Logic Models Guided Several Evaluations

Several programs provided evaluation expertise to guide and encourage program staff to evaluate their own programs. The guidance encouraged them to develop program logic models to articulate program strategy and evaluation questions. Cooperative extension has evaluation specialists in many of the state land grant universities who offer many useful evaluation tools and guidance on their Web sites. (See the Bibliography for a list of resources.) CDC provided the rationale for how the National Tobacco

Control Program addressed the policy problem (youths' smoking) and articulated the conceptual framework for how the program activities were expected to motivate people to change their behavior. CDC supports local project evaluation with financial and technical assistance and a framework for program evaluation that provides general guidance on engaging stakeholders, evaluation design, data collection and analysis, and ways to ensure that evaluation findings are used. CDC also encourages grantees to allocate about 10 percent of their program budget for program monitoring (surveillance) and evaluation.
(See www.cdc.gov/Tobacco/evaluation_manual/contents.htm).

CDC, EPA, and cooperative extension evaluation guidance all encouraged project managers to create program logic models to help articulate their program strategy and expected outcomes. Logic models characterize how a program expects to achieve its goals; they link program resources and activities to program outcomes and identify short-term and long-term outcome goals. CDC's recent evaluation guidance suggests that grantees use logic models to link inputs and activities to program outcomes and also to demonstrate how a program connects to the national and state programs. The University of Wisconsin Cooperative Extension evaluation guidance noted that local projects would find developing the program logic model to be useful in program planning, identifying measures, and explaining the program to others.

Observations

The agencies whose evaluations we studied employed a variety of strategies for evaluating their programs' effects on short-term and intermediate goals but still had difficulty assessing their contributions to long-term agency goals for social and environmental benefits. As other agencies are pressed to demonstrate the effectiveness of their information campaigns, the examples in this report might help them identify how to successfully evaluate their programs' contributions.

Several agencies drew on existing research to identify common measures; others may find that analysis of the relevant research literature can aid in designing a program evaluation. Previous research may reveal useful existing measures or clarify the expected influence of the program, as well as external factors, on its goals.

Agencies might also benefit from following the evaluation guidance that has recommended developing logic models that specify the mechanisms by which programs are expected to achieve results, as well as the specific

short-term, intermediate, and long-term outcomes they are expected to achieve.

- A logic model can help identify pertinent variables and how, when, and in whom they should be measured, as well as other factors that might affect program results. This, in turn, can help set realistic expectations about the scope of a program's likely effects. Specifying a logical trail from program activities to distant outcomes pushes program and evaluation planners to articulate the specific behavior changes and long-term outcomes they expect, thereby indicating the narrowly defined long-term outcomes that could be attributed to a program.

- Where program flexibility allows for local variation but risks losing accountability, developing a logic model can help program stakeholders talk about how diverse activities contribute to common goals and how this might be measured. Such discussion can sharpen a program's focus and can lead to the development of commonly accepted standards and measures for use across sites.

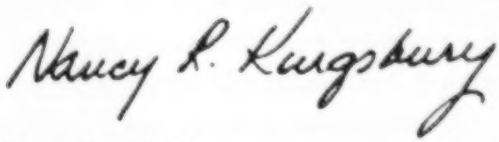
- In comprehensive initiatives that combine various approaches to achieving a goal, developing a logic model can help articulate how those approaches are intended to assist and supplement one another and can help specify how the information dissemination portion of the program is expected to contribute to their common goal. An evaluation could then assess the effects of the integrated set of efforts on the desired long-term outcomes, and it could also describe the short-term and intermediate contributions of the program's components.

Agency Comments

The agencies provided no written comments, although EPA, HHS, and USDA provided technical comments that we incorporated where appropriate throughout the report. EPA noted that the Paperwork Reduction Act requirements pose an additional challenge in effectively and efficiently measuring compliance assistance outcomes. We included this point in the discussion of follow-up surveys.

We are sending copies of this report to other relevant congressional committees and others who are interested, and we will make copies available to others on request. In addition, the report will be available at no charge on GAO's Web site at http://www.gao.gov.

If you have questions concerning this report, please call me or Stephanie Shipman at (202) 512-2700. Elaine Vaurio also made key contributions to this report.



Nancy Kingsbury
Managing Director, Applied Research
and Methods

Bibliography

Case Evaluations and Guidance

Centers for Disease Control and Prevention, Office of Smoking and Health. *Best Practices for Comprehensive Tobacco Control Programs*. Atlanta, Ga.: August 1999. http://www.cdc.gov/tobacco/bestprac.htm (September 2002).

Garet, Michael S., and others. *Designing Effective Professional Development: Lessons from the Eisenhower Program*. Document 99-3. Washington, D.C.: U.S. Department of Education, Planning and Evaluation Service, December 1999. http://www.ed.gov/inits/teachers/eisenhower/ (September 2002).

Hornik, Robert, and others. *Evaluation of the National Youth Anti-Drug Media Campaign: Historical Trends in Drug Use and Design of the Phase III Evaluation*. Prepared for the National Institute on Drug Abuse. Rockville, Md.: Westat, July 2000. http://www.whitehousedrugpolicy.gov/publications (September 2002).

Hornik, Robert, and others. *Evaluation of the National Youth Anti-Drug Media Campaign: Third Semi-Annual Report of Findings*. Prepared for the National Institute on Drug Abuse. Rockville, Md.: Westat, October 2001. http://www.mediacampaign.org/publications/index.html (September 2002).

Kiernan, Nancy Ellen. "Reduce Bias with Retrospective Questions." Penn State University Cooperative Extension Tipsheet 30, University Park, Pennsylvania, 2001. http://www.extension.psu.edu/evaluation/ (September 2002).

Kiernan, Nancy Ellen. "Using Observation to Evaluate Skills." Penn State University Cooperative Extension Tipsheet 61, University Park, Pennsylvania, 2001. http://www.extension.psu.edu/evaluation/ (September 2002).

MacDonald, Goldie, and others. *Introduction to Program Evaluation for Comprehensive Tobacco Control Programs*. Atlanta, Ga.: Centers for Disease Control and Prevention, November 2001. http://www.cdc.gov/tobacco/evaluation_manual/contents.htm (September 2002).

Office of National Drug Control Policy. *The National Youth Anti-Drug Media Campaign: Communications Strategy Statement.* Washington, D.C.: Executive Office of the President, n.d. http://www.mediacampaign.org/publications/index.html (September 2002).

Ohio State University Cooperative Extension. Program Development and Evaluation. http://www.ag.ohio-state.edu/~pde/ (September 2002).

Penn State University. College of Agricultural Sciences, Cooperative Extension and Outreach, Program Evaluation http://www.extension.psu.edu/evaluation/ (September 2002).

Porter, Andrew C., and others. *Does Professional Development Change Teaching Practice? Results from a Three-Year Study.* Document 2000-04. Washington, D.C.: U.S. Department of Education, Office of the Under Secretary, October 2000. http://www.ed.gov/offices/OUS/PES/school_improvement.html#subepdp2 (September 2002).

Rockwell, S. Kay, and Harriet Kohn. "Post-Then-Pre Evaluation." *Journal of Extension* 27:2 (summer 1989). http://www.joe.org/joe/1989summer/a5.html (September 2002).

Taylor-Powell, Ellen. "The Logic Model: A Program Performance Framework." University of Wisconsin Cooperative Extension, Madison, Wisconsin, 62 pages, n.d. http://www.uwex.edu/ces/pdande (September 2002).

Taylor-Powell, Ellen, and Marcus Renner. "Collecting Evaluation Data: End-of-Session Questionnaires." University of Wisconsin Cooperative Extension document G3658-11, Madison, Wisconsin, September 2000. http://www.uwex.edu/ces/pdande (September 2002).

Taylor-Powell, Ellen, and Sara Steele. "Collecting Evaluation Data: Direct Observation." University of Wisconsin Cooperative Extension document G3658-5, Madison, Wisconsin, 1996. http://www.uwex.edu/ces/pdande/evaluation/evaldocs.html (September 2002).

U.S. Department of Agriculture, Expanded Food and Nutrition Education Program. *EFNEP 2001 Program Impacts Booklet.* Washington, D.C.: June 2002. http://www.reeusda.gov/f4hn/efnep/factsheet.htm (September 2002).

U.S. Department of Agriculture, Expanded Food and Nutrition Education Program. ERS4 (Evaluation/Reporting System). Washington, D.C.: April 9, 2001. http://www.reeusda.gov/ers4/home.htm (September 2002).

U.S. Department of Agriculture, Expanded Food and Nutrition Education Program. *Virginia EFNEP Cost Benefit Analysis*. Fact Sheet. Washington, D.C.: n.d. http://www.reeusda.gov/f4hn/efnep.htm (September 2002).

U.S. Environmental Protection Agency, Office of Enforcement and Compliance Assurance. *Guide for Measuring Compliance Assistance Outcomes*. EPA300-B-02-011. Washington, D.C.: June 2002. http://www.epa.gov/compliance/planning/results/tools.html (September 2002).

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. "Framework for Program Evaluation in Public Health." *Morbidity and Mortality Weekly Report* 48:RR-11 (1999). (September 2002).

University of Wisconsin Cooperative Extension. Program Development and Evaluation, Evaluation. http://www.uwex.edu/ces/pdande/evaluation/index.htm (September 2002).

Other Evaluation Guidance and Tools

American Evaluation Association, Extension Education Evaluation Topical Interest Group http://www.danr.ucop.edu/eee-aea/ (September 2002).

CYFERnet. Children, Youth, and Families Education and Research Network. Evaluation Resources http://twosocks.ces.ncsu.edu/cyfdb/browse_2.php?search=Evaluation (September 2002).

Schwarz, Norbert, and Daphna Oyserman. "Asking Questions about Behavior: Cognition, Communication, and Questionnaire Construction." *American Journal of Evaluation* 22:2 (summer 2001): 127–60.

Southern Regional Program and Staff Development Committee. "Evaluation and Accountability Resources: A Collaboration Project of the Southern Region Program and Staff Development Committee." Kentucky Cooperative Extension Service. http://www.ca.uky.edu/agpsd/soregion.htm (September 2002).

Related GAO Products

Program Evaluation: Studies Helped Agencies Measure or Explain Program Performance. GAO/GGD-00-204. Washington, D.C.: September 29, 2000.

Anti-Drug Media Campaign: ONDCP Met Most Mandates, but Evaluations of Impact Are Inconclusive. GAO/GGD/HEHS-00-153. Washington, D.C.: July 31, 2000.

Managing for Results: Measuring Program Results That Are under Limited Federal Control. GAO/GGD-99-16. Washington, D.C.: December 11, 1998.

Grant Programs: Design Features Shape Flexibility, Accountability, and Performance Information. GAO/GGD-98-137. Washington, D.C.: June 22, 1998.

Program Evaluation: Agencies Challenged by New Demand for Information on Program Results. GAO/GGD-98-53. Washington, D.C.: April 24, 1998.

Managing for Results: Analytic Challenges in Measuring Performance. GAO/HHS/GGD-97-138. Washington, D.C.: May 30, 1997.

Program Evaluation: Improving the Flow of Information to the Congress. GAO/PEMD-95-1. Washington, D.C.: January 30, 1995.

Designing Evaluations. GAO/PEMD-10.1.4. Washington, D.C.: May 1991.

GAO's Mission

The General Accounting Office, the investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through the Internet. GAO's Web site (www.gao.gov) contains abstracts and full-text files of current reports and testimony and an expanding archive of older products. The Web site features a search engine to help you locate documents using key words and phrases. You can print these documents in their entirety, including charts and other graphics.

Each day, GAO issues a list of newly released reports, testimony, and correspondence. GAO posts this list, known as "Today's Reports," on its Web site daily. The list contains links to the full-text document files. To have GAO e-mail this list to you every afternoon, go to www.gao.gov and select "Subscribe to daily E-mail alert for newly released products" under the GAO Reports heading.

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. General Accounting Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone: Voice: (202) 512-6000
TDD: (202) 512-2537
Fax: (202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm
E-mail: fraudnet@gao.gov
Automated answering system: (800) 424-5454 or (202) 512-7470

Public Affairs

Jeff Nelligan, managing director, NelliganJ@gao.gov (202) 512-4800
U.S. General Accounting Office, 441 G Street NW, Room 7149
Washington, D.C. 20548

PRINTED ON RECYCLED PAPER

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
X	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").